PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-056693
(43)Date of publication of application : 03.03.1995

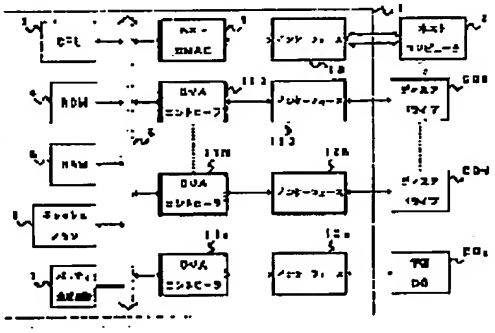
(51)Int.Cl. G06F 3/06
G06F 3/06
G06F 12/08

(21)Application number : 05-201676
(22)Date of filing : 13.08.1993

(71)Applicant : TOSHIBA CORP
SORD COMPUT CORP
(72)Inventor : SEGAWA KIYOSHI

(54) DISK ARRAY DEVICE AND DISK WRITING CONTROLLER

(57)Abstract:
PURPOSE: To attain the efficiency and high speed of the write access of data whose size is especially small by transferring write data to a preliminary disk drive, and storing the data in it before executing a write back processing at the time of the write access.
CONSTITUTION: The write cache of storing the write data from a host computer 2 in a cache memory 6 is executed. Afterwards, the write back processing of transferring the write data from the cache memory 6 to each disk drive DD0–DDN is executed. Before the write back processing is executed, the write data are transferred to and stored in a preliminary disk driver DDS by a preliminary disk DMA controller 11S. Also, at the time of the write back processing, when the transfer of the write data from memory 6 to each disk driver DD 0–DDN is imppossible, the write data stored in the preliminary disk drive DDS are read and used.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平7－56693

(43)公開日　平成７年(1995)３月３日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ	3/06	５４０	7165－5B		
		３０１　Ｊ			
	12/08	３２０	7608－5B		

審査請求　未請求　請求項の数４　ＯＬ　（全 12 頁）

(21)出願番号　特願平5－201676

(22)出願日　平成５年(1993)８月13日

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地

(71)出願人　000108362
ソード株式会社
千葉県千葉市美浜区真砂５丁目20番７号

(72)発明者　瀬川　清
千葉県千葉市真砂５丁目20番７号　ソード株式会社内

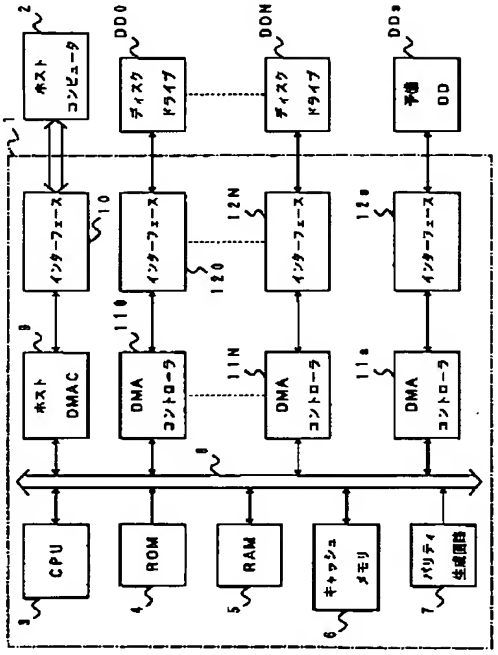
(74)代理人　弁理士　鈴江　武彦

(54)【発明の名称】　ディスク・アレイ装置とディスク書込み制御装置

(57)【要約】

【目的】本発明の目的は、ライトキャッシュ方式を採用したディスク・アレイ装置において、コストの増大化やシステム構成の複雑化を招くことなく、特に小さいサイズのデータのライトアクセスの高速かつ効率化を図ると共に、データを確実に保護して装置の高信頼性を実現することにある。

【構成】ホストコンピュータ２からのライトデータはキャッシュメモリ６に格納するライトキャッシュが実行される。この後に、キャッシュメモリ６から各ディスクドライブＤＤ０～ＤＤＮにライトデータを転送するライトバック処理が実行される。このライトバック処理の実行前に、予備ディスクＤＭＡＣ１１ｓはライトデータを予備ディスクドライブＤＤｓに転送して格納する。ＣＰＵ３は、ライトバック処理が不可能な時に、キャッシュメモリ手段から各ディスクドライブへのライトデータの転送が不可能なときに、予備ディスクドライブＤＤｓに格納したライトデータを読出して使用するバックアップ処理を実行する。



1

【特許請求の範囲】

【請求項１】　予備ディスクドライブを含む複数のディスクドライブと、
外部から転送されたライトデータを格納するキャッシュメモリ手段と、
このキャッシュメモリ手段から前記予備ディスクドライブを除く前記各ディスクドライブに前記ライトデータを転送するライトバック処理の実行前に、前記ライトデータを前記予備ディスクドライブに転送して格納する予備ディスク制御手段とを具備したことを特徴とするディスク・アレイ装置。

【請求項２】　予備ディスクドライブを含む複数のディスクドライブと、
外部から転送されたライトデータを格納するキャッシュメモリ手段と、
このキャッシュメモリ手段から前記予備ディスクドライブを除く前記各ディスクドライブに前記ライトデータを転送するライトバック処理の実行前に、前記ライトデータを前記予備ディスクドライブに転送して格納する予備ディスク制御手段と、
前記ライトバック処理時に、前記キャッシュメモリ手段からの前記各ディスクドライブへの前記ライトデータの転送が不可能なときに、前記予備ディスクドライブに格納した前記ライトデータを読出して使用するバックアップ制御手段とを具備したことを特徴とするディスク・アレイ装置。

【請求項３】　予備ディスクドライブを含む複数のディスクドライブと、
外部から転送されたライトデータを格納するキャッシュメモリ手段と、
前記ライトデータの転送データ量に基づいて前記ライトデータのサイズが相対的に小さい場合に、前記予備ディスクドライブをバックアップとして使用するモードを決定するモード決定手段と、
前記モード時に、前記キャッシュメモリ手段から前記予備ディスクドライブを除く前記各ディスクドライブに前記ライトデータを転送するライトバック処理の実行前に、前記ライトデータを前記予備ディスクドライブに転送して格納する予備ディスク制御手段とを具備したことを特徴とするディスク・アレイ装置。

【請求項４】　予備ディスクドライブを含む複数のディスクドライブを有するディスク・システムにおいて、
外部から転送されたライトデータを格納するキャッシュメモリ手段と、
このキャッシュメモリ手段に格納された前記ライトデータを前記予備ディスクドライブに格納するバックアップ制御時に、前記ライトデータの識別情報、前記ライトデータを代用する無効データおよび前記ライトデータを前記予備ディスクドライブを除く前記各ディスクドライブに格納したことを指示する指示情報のそれぞれを生成するデータ生成手段と、
前記バックアップ制御時に、前記予備ディスクドライブの連続するアドレスに前記無効データ、前記識別情報、前記ライトデータおよび前記指示情報を逐次書込むデータ書込み制御手段とを具備したことを特徴とするディスク書込み制御装置。

2

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、複数のディスクドライブを有し、ホストコンピュータからのデータを各ディスクドライブに分散して格納するディスク・アレイ装置に関する。

【０００２】

【従来の技術】従来、コンピュータシステムに使用される外部記憶装置として、ＲＡＩＤ（Ｒｅｄｕｎｄａｎｔ　Ａｒｒａｙｓ　ｏｆ　Ｉｎｅｘｐｅｎｓｉｖｅ　Ｄｉｓｋｓ）とも呼ばれているディスク・アレイ装置が周知である。ディスク・アレイ装置は、複数のディスクドライブ（ハードディスク装置）を有し、ホストコンピュータから転送されたデータを各ディスクドライブに分散して格納する。このため、ホストコンピュータからは、あたかも１台のハードディスク装置をアクセスするように見える。

【０００３】ディスク・アレイ装置には、レベル１～５の５段階に分けられた異なる方式が開発されている。例えばレベル３の方式は、データをビット単位のインターリーピング（ｉｎｔｅｒｌｅａｖｉｎｇ）により各ディスクドライブに分散して格納する。また、レベル４の方式は、ホストコンピュータからのデータを分散するだけでなく、エラー訂正用のパリティデータを格納する専用のディスクドライブを備えた方式である。

【０００４】さらに、レベル５の方式はレベル４と同様に、データをセクタ単位のインターリーピングにより、各ディスクドライブに分散して格納する。レベル５とレベル４の各方式の相違は、レベル４がパリティデータの専用ディスクドライブを備えているに対して、レベル５がパリティデータを各ディスクドライブに分散して格納する点にある。

【０００５】即ち、レベル５の方式は、図８に示すように、複数のディスクドライブＤＤ０～ＤＤ３にデータを論理セクタアドレス０～１１に分散して格納すると共に、パリティデータＰＤ１～ＰＤ４も分散して格納する。

【０００６】ここで、１セクタ（論理セクタアドレス）には通常では５１２バイトのデータが格納される。ディスクドライブＤＤ０～ＤＤ３では、同一物理セクタＮ（Ｎ＝０～３）に属する４つの論理アドレスセクタは１つのパリティグループを構成する。即ち、パリティデータＰＤ１～ＰＤ４はそれぞれ、パリティグループを構成する各論理アドレスセクタの各データと同一バイト位置

に書込まれる。例えばパリティデータＰＤ１は、論理アドレスセクタ０～２とパリティグループを構成する。

【０００７】このようなレベル５の方式では、１台のディスクドライブ（例えばＤＤ０）が故障しても、残りの３台のディスクドライブ（ＤＤ１～ＤＤ２）のデータとパリティデータＰＤ１を利用することにより、故障したディスクドライブ（例えばＤＤ０）のデータを修復することができる。したがって、単独のハードディスク装置と比較して、高い信頼性を確保することができる。

【０００８】しかしながら、２台以上のディスクドライブに故障が発生すると、前記の方法ではデータの修復は不可能である。このため、図８に示すように、予備ディスクドライブＤＤｓを用意して、１台目のディスクドライブの故障が発生したときに、前記の方法でデータを修復し、この修復データを予備ディスクドライブＤＤｓにコピーする方式が開発されている。この方式では、当然ながら予備ディスクドライブＤＤｓは、１台目のディスクドライブに故障が発生するまでアクセスされない。

【０００９】一方、レベル５の方式はレベル３の方式と比較して、画像データ等の転送データ量（データのサイズ）が大きいデータよりも、小さいサイズのデータのアクセスに有効である。レベル３の方式は、１セクタ分のデータをアクセスするときに、複数のディスクドライブの全てをアクセスする必要がある。このため、データ転送速度は高速となるため大きいサイズのデータ転送には都合がいいが、小さい単位のデータアクセスでは効率が低下する。

【００１０】しかし、レベル５においても、小さいサイズのデータをライトアクセスするときには、パリティデータの書換えを必要とするため性能が低下する。このため、特にライトアクセス時に、キャッシュメモリを使用するライトキャッシュ方式を採用した装置が開発されている。即ち、例えば１セクタ分のライトアクセスでは、書換える前のデータ（旧ユーザデータ）とそれに対応するパリティデータ（旧パリティデータ）をそれぞれリードアクセスする。これらの旧ユーザデータ、旧パリティデータおよび新たなライトデータ（新ユーザデータ）から新たなパリティデータ（新パリティデータ）を生成する。したがって、１セクタ分のライトアクセスには、２リードアクセス、新ユーザデータのライトアクセスおよび新パリティデータのライトアクセスの計４アクセスが必要となる。

【００１１】このような欠点を解消するために、レベル５に前記のライトキャッシュ方式を採用した装置がある。この方式は、ホストコンピュータからライトデータ（新ユーザデータ）を受信すると、キャッシュメモリに格納してライトコマンドを終了する。この後に、ディスクドライブのアイドル時（非アクセス時）に、ライトデータと新パリティデータを指定の各セクタに書込むライトバック処理を実行する。

【００１２】

【発明が解決しようとする課題】従来のレベル５のディスク・アレイ装置において、ライトキャッシュ方式を採用することにより、小さいサイズのデータのライトアクセスを高速かつ効率的に行なうことが可能となる。しかしながら、この方式では、キャッシュメモリからデータをディスクドライブに転送するライトバック処理が必要となる。このライトバック処理の実行終了前に、キャッシュメモリに故障が発生したり、またはキャッシュメモリの電源供給が停止して、キャッシュメモリに格納されたデータが破壊または消去するような事態が発生する可能性がある。このような事態が発生すれば、ライトバック処理は不可能となる。このため、単にライトキャッシュを採用した方式では、データ保護が不完全であり、装置の信頼性が低いという問題がある。

【００１３】信頼性を高めるためには、キャッシュメモリやディスク・アレイ・コントローラの２重化、またはキャッシュメモリに不揮発性メモリや無停電電源装置を使用するなどの方法が考えられる。しかし、このような方法を採用すると、装置全体のコストが増大し、かつシステム構成の複雑化を招く欠点がある。

【００１４】本発明の目的は、ライトキャッシュ方式を採用したディスク・アレイ装置において、特に小さいサイズのデータのライトアクセスの高速かつ効率化を図ると共に、コストの増大化やシステム構成の複雑化を招くことなく、データを確実に保護して装置の高信頼性を実現することにある。

【００１５】

【課題を解決するための手段】本発明は、例えばレベル５のディスク・アレイ装置において、予備ディスクドライブを含む複数のディスクドライブ、ライトデータを格納するキャッシュメモリ手段およびライトアクセス時のライトバック処理の実行前に、ライトデータを予備ディスクドライブに転送して格納する予備ディスク制御手段を有する装置である。

【００１６】さらに、本発明はライトバック処理時に、キャッシュメモリ手段から各ディスクドライブへのライトデータの転送が不可能なときに、予備ディスクドライブに格納したライトデータを読出して使用するバックアップ制御手段を有する装置である。

【００１７】

【作用】本発明では、ホストコンピュータからのライトデータはキャッシュメモリ手段に格納するライトキャッシュが実行される。この後に、キャッシュメモリ手段から各ディスクドライブにライトデータを転送するライトバック処理が実行される。このライトバック処理の実行前に、予備ディスク制御手段はライトデータを予備ディスクドライブに転送して格納する。

【００１８】さらに、本発明では、バックアップ制御手段はライトバック処理時に、キャッシュメモリ手段から

5

各ディスクドライブへのライトデータの転送が不可能なときに、予備ディスクドライブに格納したライトデータを読出して使用する。

【００１９】

【実施例】以下図面を参照して本発明の実施例を説明する。図１は同実施例に係わるディスク・アレイ装置（ＲＡＩＤ装置）の要部を示すブロック図、図２乃至図４および図７は同実施例の動作を説明するためのフローチャート、図５および図６は同実施例に係わる予備ディスクドライブの記憶内容を説明するための概念図である。

【００２０】本装置は、図１に示すように、複数のディスクドライブＤＤ０～ＤＤＮと１台の予備ディスクドライブＤＤｓ、および各ディスクドライブを制御するディスク・アレイ・コントローラ（以下単にコントローラと称する）１を有する。コントローラ１は、レベル５の方式により各ディスクドライブＤＤ０～ＤＤＮを制御する。即ち、コントローラ１は、ホストコンピュータ２から転送されるライトデータを、セクタ単位のインターリービングにより各ディスクドライブＤＤ０～ＤＤＮに分散して格納する。さらに、コントローラ１は、ライトデータに対応するパリティデータを生成して各ディスクドライブＤＤ０～ＤＤＮに分散して格納する。ここで、ホストコンピュータ２との間で転送するリードデータまたはライトデータを、パリティデータと区別する場合にはユーザデータと称する。

【００２１】コントローラ１は、マイクロプロセッサ（ＣＰＵ）３、読出し専用メモリ（ＲＯＭ）４、リード／ライトメモリ（ＲＡＭ）５、キャッシュメモリ６、パリティ生成回路７および内部バス８を有する。ＣＰＵ３は、ＲＯＭ４に予め格納されたプログラムにより、各ディスクドライブＤＤ０～ＤＤＮのリード／ライト制御、本発明に係わるライトバック処理およびバックアップ処理を実行する。ＲＡＭ５はＣＰＵ３のワークメモリとして使用される。キャッシュメモリ６は、リード／ライトデータを一時的に格納するための高速バッファメモリである。パリティ生成回路７は、ホストコンピュータ２からのライトデータに対応するパリティデータを生成するための回路である。

【００２２】さらに、コントローラ１は、ホストＤＭＡコントローラ（ホストＤＭＡＣ）９、インターフェース１０、ディスクＤＭＡコントローラ（ディスクＤＭＡＣ）１１０～１１Ｎ、インターフェース１２０～１２Ｎ、予備ディスクＤＭＡコントローラ（予備ＤＭＡＣ）１１ｓおよびインターフェース１２ｓを有する。ホストＤＭＡＣ９は、ホストバス１３と内部バス８間のデータ転送を制御し、本発明に関係するライトアクセス時にホストバス１３からのライトデータをキャッシュメモリ６に転送する。インターフェース１０はホストバス１３に接続されて、ホストコンピュータ２とのデータ交換を実行する。

6

【００２３】ディスクＤＭＡＣ１１０～１１Ｎはそれぞれのインターフェース１２０～１２Ｎを介して、各ディスクドライブＤＤ０～ＤＤＮと内部バス８との間のデータ転送を制御する。予備ＤＭＡＣ１１ｓはインターフェース１２ｓを介して、予備ディスクドライブＤＤｓと内部バス８との間のデータ転送を制御する。

【００２４】次に、同実施例の動作を説明する。

（基本動作）まず、図２のフローチャートを参照して、レベル５のディスク・アレイ装置の基本的動作を説明する。ホストコンピュータ（ホストＣＰＵ）２からリード／ライト（Ｒ／Ｗ）コマンドがホストバス１３を介して転送されると、ホストＤＭＡＣ９は内部バス８を通じてＲ／ＷコマンドをＲＡＭ５に転送する（ステップＳ１，Ｓ２）。ＣＰＵ３は、ホストＣＰＵ２からのＲ／Ｗコマンドを解読してディスクコマンドに変換する（ステップＳ３）。ディスクコマンドは、各ディスクドライブＤＤ０～ＤＤＮのリード／ライト制御を実行するためのコマンドである。

【００２５】ＣＰＵ３は変換したディスクコマンドをＲＡＭ５に格納する。ディスクＤＭＡＣ１１０～１１Ｎは、ＲＡＭ５からリード／ライト動作を指示するディスクコマンドを各ディスクドライブＤＤ０～ＤＤＮに転送する。

（リードアクセス）ディスクコマンドがリードコマンドであれば（ステップＳ４のＹＥＳ）、ディスクＤＭＡＣ１１０～１１Ｎはリードコマンドを各ディスクドライブＤＤ０～ＤＤＮに転送する（ステップＳ６）。各ディスクドライブＤＤ０～ＤＤＮは、指定された論理セクタアドレスのデータを読出すリードアクセスを実行する（ステップＳ７）。ディスクＤＭＡＣ１１０～１１Ｎは、読出したリードデータをキャッシュメモリ６に転送する（ステップＳ８）。

【００２６】ホストＤＭＡＣ９は、転送順序をチェックしながら、キャッシュメモリ６に格納されたリードデータをインターフェース１０を介してホストＣＰＵ２に転送する（ステップＳ９）。ここで、例えばディスクドライブＤＤ０のディスクにディフェクトが存在したり、データの破壊が発生したりして、データの読出し動作が不可能な場合には、ＣＰＵ３はデータの修復処理を実行する（ステップＳ１０のＮＯ，Ｓ１１）。即ち、ＣＰＵ３は、ディスクＤＭＡＣ１１０を介して、リードできないデータの同一パリティグループに属するユーザデータとパリティデータを読出し、パリティ計算に基づいて元のデータを修復してキャッシュメモリ６に格納する。ホストＤＭＡＣ９は、キャッシュメモリ６から修復したデータをリードデータとしてホストＣＰＵ２に転送する。ＣＰＵ３は、正常にリードコマンドが終了すると、正常終了ステータスをホストＣＰＵ２に送信する。

【００２７】ＣＰＵ３は、前記のように例えばディスクドライブＤＤ０に障害が発生すると、ディフェクトの交

替処理を実行したり、または故障したディスクドライブＤＤ０の代わりに予備ディスクドライブＤＤｓを代替する。この代替処理では、故障したディスクドライブＤＤ０に格納した全データを修復して、予備ディスクドライブＤＤｓに格納する。以後、ＣＰＵ３は予備ディスクドライブＤＤｓを正規のディスクドライブＤＤ０として使用する。

【００２８】ホストＣＰＵ２からのコマンドがライトコマンドであれば（ステップＳ４のＮＯ）、ＣＰＵ３は各ディスクドライブＤＤ０～ＤＤＮに対するライトアクセスを実行する（ステップＳ５）。同実施例では、ホストＤＭＡＣ９は、ホストＣＰＵ２からのライトデータをキャッシュメモリ６に格納するライトキャッシュを実行する。

【００２９】以下図３のフローチャートを参照して、同実施例に係わるライトアクセスの動作を説明する。同実施例では、ＣＰＵ３はライトデータのサイズを判定し、比較的大きいサイズのライトデータあれば、通常のライトアクセス動作を実行する。また、比較的小さいサイズのライトデータあれば、本発明に係わるライトアクセス動作を実行する。

【００３０】即ち、ＣＰＵ３は、ホストＤＭＡＣ９により転送されたライトコマンドにより、ライトデータの転送データ量を認識して、ライトデータのサイズを判定する（ステップＳ１２）。このとき、ＣＰＵ３はキャッシュメモリ６の使用可能な格納容量も検出する。

（大きいサイズのライトアクセス）ライトデータが例えば画像データのように大きいサイズのデータであれば（ステップＳ１３のＮＯ）、ホストＤＭＡＣ９は、ホストＣＰＵ２からのライトデータをキャッシュメモリ６に転送する（ステップＳ１４）。ＣＰＵ３は、パリティ生成回路７を使用して、ライトデータに対応するパリティデータを生成してキャッシュメモリ６に格納する（ステップＳ１５）。即ち、ＣＰＵ３は、ライトデータの書換え対象である旧ユーザデータと旧パリティデータを読出し、これらの旧データに基づいて新たなパリティデータを生成する。この場合、キャッシュメモリ６は単なるバッファメモリとして使用される。

【００３１】ディスクＤＭＡＣ１１０～１１Ｎは、キャッシュメモリ６からライトデータと新パリティデータを、各ディスクドライブＤＤ０～ＤＤＮに転送する。各ディスクドライブＤＤ０～ＤＤＮは、ライトデータと新パリティデータを、ディスクコマンドにより指定された論理セクタアドレスに格納する（ステップＳ１７）。ＣＰＵ３は、各ディスクドライブＤＤ０～ＤＤＮの全てが正常にライトコマンドを終了すると、正常終了ステータスをホストＣＰＵ２を送信する。

【００３２】このようなライトアクセス時においても、ディスクドライブＤＤ０～ＤＤＮに障害が発生すると、前記のリードアクセスの場合と同様に、ＣＰＵ３はディスクドライブの代わりに予備ディスクドライブＤＤｓを代替する。この代替処理では、故障したディスクドライブＤＤ０に格納した全データを修復して、予備ディスクドライブＤＤｓに格納する。以後、ＣＰＵ３は予備ディスクドライブＤＤｓを正規のディスクドライブとして使用する。

フェクトの交替処理を実行したり、または故障したディスクドライブの代わりに予備ディスクドライブＤＤｓを代替する。この代替処理では、故障したディスクドライブＤＤ０に格納した全データを修復して、予備ディスクドライブＤＤｓに格納する。以後、ＣＰＵ３は予備ディスクドライブＤＤｓを正規のディスクドライブとして使用する。

（小さいサイズのライトアクセス）ライトデータが小さいサイズのデータであれば（ステップＳ１３のＹＥＳ）、ホストＤＭＡＣ９は、ホストＣＰＵ２からのライトデータをキャッシュメモリ６に転送する（ステップＳ１８）。次に、本発明では、予備ＤＭＡＣ１１ｓはインターフェース１２ｓを介して、キャッシュメモリ６に格納されたライトデータを予備ディスクドライブＤＤｓに転送する（ステップＳ１９）。これにより、予備ディスクドライブＤＤｓはライトデータを格納し、バックアップすることになる（ステップＳ２０）。このとき、予備ＤＭＡＣ１１ｓは、ホストＤＭＡＣ９により転送されるライトデータをそのまま予備ディスクドライブＤＤｓに転送してもよい。

【００３３】この後、ＣＰＵ３は、各ディスクドライブＤＤ０～ＤＤＮのアイドル時（非アクセス時）に、バックグラウンドジョブにより、キャッシュメモリ６に格納されたライトデータを各ディスクドライブＤＤ０～ＤＤＮに格納するライトバック処理を実行する（ステップＳ２１）。

【００３４】ライトバック処理では、図４に示すように、ＣＰＵ３は、パリティ生成回路７を使用して、ライトデータに対応するパリティデータを生成してキャッシュメモリ６に格納する（ステップＳ２２）。ディスクＤＭＡＣ１１０～１１Ｎは、キャッシュメモリ６からライトデータと新パリティデータを、各ディスクドライブＤＤ０～ＤＤＮに転送する（ステップＳ２３）。各ディスクドライブＤＤ０～ＤＤＮは、ライトデータと新パリティデータを、ディスクコマンドにより指定された論理セクタアドレスに格納する（ステップＳ２４）。

【００３５】ＣＰＵ３は、各ディスクドライブＤＤ０～ＤＤＮの全てが正常にライトコマンドを終了すると、正常終了ステータスをホストＣＰＵ２を送信する（ステップＳ２５のＹＥＳ）。このようなライトバック処理が正常に終了するまで、予備ディスクドライブＤＤｓにはライトデータがバックアップされている。

【００３６】ここで、ライトバック処理が正常に終了する前に、キャッシュメモリ６に故障が発生したり、またはキャッシュメモリ６の電源供給が停止して、キャッシュメモリ６に格納されたデータが破壊または消去するような事態が発生したとする（ステップＳ２５のＮＯ）。予備ＤＭＡＣ１１ｓは、ＣＰＵ３の指示に応じて、予備ディスクドライブＤＤｓにバックアップしてあるライトデータをリードする（ステップＳ２６）。ディスクＤＭ

ＡＣ１１０～１１Ｎは、予備ディスクドライブＤＤｓからリードされたライトデータを、指定された論理セクタアドレスにライトバックする（ステップＳ２７）。このとき、ＣＰＵ３は、パリティ生成回路７を使用して、ライトデータに対応するパリティデータを新たに生成する。

【００３７】このようにして、小さいサイズのライトデータを書込む場合には、キャッシュメモリ６に格納すると共に、予備ディスクドライブＤＤｓにバックアップする。このライトキャッシュとバックアップ処理により、ライトコマンドは終了となる。この後、ホストＣＰＵ２からのリード／ライトアクセスのない各ディスクドライブＤＤ０～ＤＤＮのアイドル時に、キャッシュメモリ６から各ディスクドライブＤＤ０～ＤＤＮに対して、ライトデータのライトバック処理が実行される。

【００３８】予備ディスクドライブＤＤｓにバックアップされたライトデータは、ライトバック処理が終了するまで保存されている。ライトバック処理が終了する前に、キャッシュメモリ６に故障または電源供給の停止の事態が発生して、ライトバック処理が不可能になったときに、予備ディスクドライブＤＤｓにバックアップされたライトデータが利用される。したがって、結果的にライトデータは常に確実に保護されることになり、装置の信頼性を高めることが可能となる。

【００３９】ところで、ディスクドライブＤＤ０～ＤＤＮに障害が発生した場合には、前記の大きいサイズのライトデータのアクセス時と同様に、予備ディスクドライブＤＤｓを利用してデータの修復処理を実行する。このとき、まずキャッシュメモリ６に格納されたライトデータの全てをライトバック処理する。この後に、故障したディスクドライブに格納した全データを修復し、予備ディスクドライブＤＤｓに格納する。以後、ライトキャッシュを禁止して、キャッシュメモリ６を単なるバッファメモリとして使用し、ＣＰＵ３は予備ディスクドライブＤＤｓを正規のディスクドライブとして使用する。

（予備ディスクドライブＤＤｓの動作）次に、本発明に係わる予備ディスクドライブＤＤｓの動作を、図５乃至図７を参照して具体的に説明する。

【００４０】まず、ＣＰＵ３は予備ディスクドライブＤＤｓをライトアクセスするときに、シーケンシャル・ライトコマンドを出力する。即ち、図５に示すように、アクセスされるセクタアドレスは初期値（アドレス０）からシーケンシャルに増加し、順次セクタ単位のブロックデータが各アドレスに格納される。同実施例では、キャッシュメモリ６に格納されたライトデータが全て各ディスクドライブＤＤ０～ＤＤＮにライトバックされたときに、予備ディスクドライブＤＤｓのセクタアドレスは初期値に復帰する。また、１回のライトコマンドにより、ライトできるブロック数はコマンド，ステータスのやりとりによるオーバヘッドを十分に無視できる程度の大きい値とする。

【００４１】図７のフローチャートに示すように、ＣＰＵ３は、シーケンシャル・ライトコマンドを出力して、予備ディスクドライブＤＤｓにバックアップすべきライトデータをライトする制御を行なう（ステップＳ３０）。ここで、前記の大きいサイズのライトデータのライトアクセスの場合には、バックアップすべきライトデータはないため、ＣＰＵ３はヌルブロック（ダミーデータ）をＲＡＭ５に生成する（ステップＳ３１のＮＯ，Ｓ３２）。予備ＤＭＡＣ１１ｓは、ＣＰＵ３により生成されたヌルブロックを転送し、図５に示すように、例えばセクタアドレス０にヌルブロックを格納する（ステップＳ３３）。

【００４２】一方、ホストＣＰＵ２からバックアップすべき小さいサイズのライトデータが転送されると、前記のように、ライトデータはキャッシュメモリ６に格納されて、ライトバック処理が実行される（ステップＳ３１のＹＥＳ，Ｓ３４）。ＣＰＵ３は、１ブロック長のヘッダブロックをＲＡＭ５に生成する（ステップＳ３５）。予備ＤＭＡＣ１１ｓは、ＲＡＭ５から予備ディスクドライブＤＤｓにヘッダブロックを転送し、例えばセクタアドレス２に格納する（ステップＳ３６）。続いて、予備ＤＭＡＣ１１ｓは、キャッシュメモリ６から予備ディスクドライブＤＤｓにライトデータ（ユーザデータ）を転送し、例えばセクタアドレス３～６に格納する（ステップＳ３７）。

【００４３】ここで、予備ＤＭＡＣ１１ｓは、ライトデータの転送時に予備ディスクドライブＤＤｓのディスク回転待ちの状態が発生しないように、ヌルブロックとヘッダブロックを連続的に転送する。ＣＰＵ３は、ライトデータの転送後にヌルブロックを転送し、ライトバック処理に移行する（ステップＳ３８，Ｓ３９のＹＥＳ）。即ち、キャッシュメモリ６に格納されたライトデータは、各ディスクドライブＤＤ０～ＤＤＮにライトバックされる（ステップＳ４０）。ライトデータがライトバックされると、ＣＰＵ３はコンプリートブロック（Ｃブロック）をＲＡＭ５に生成する（ステップＳ４１）。予備ＤＭＡＣ１１ｓは、ＲＡＭ５から予備ディスクドライブＤＤｓにコンプリートブロックを転送し、例えばセクタアドレス８に格納する（ステップＳ４２）。

【００４４】ここで、ヌルブロック、ヘッダブロックおよびＣブロックは、図６に示すように、それぞれ５１２バイトのデータ長からなる。ヌルブロックは、１バイトのヌルコード（例えば００ｈ）と１バイトのバックアップＩＤコード（以下単にＩＤと称する）を有する。残りのバイトは無効データからなる。ヌルコードは、そのブロック（セクタ）にはユーザデータが存在しないことを示すコードである。

【００４５】ヘッダブロックは、１バイトのヘッダコード（例えば０１ｈ）、１バイトのＩＤ、１バイトのコマ

ンドタグおよびホストＣＰＵ２から転送されたライトコマンドを有する。ヘッダコードは、ユーザデータの先頭部であることを示すコードである。Ｃブロックは、１バイトのコンプリートコード（例えば０２ｈ）、１バイトのＩＤおよび１バイトのコマンドタグを有する。ＩＤとコマンドタグはライトバック処理が不可能な場合に、ライトバックできなかったユーザデータ（ライトデータ）を検索するために使用される。ＩＤは、予備ディスクドライブＤＤｓにシーケンシャル・ライトが実行されて、アドレスが初期値に復帰したときに「１」だけインクリメントされる。ライトバックされなかったユーザデータを検索するときは、予備ディスクドライブＤＤｓのセクタアドレス（初期値から）をシーケンシャルにリードして、初期値のアドレスの内容（ヌルブロック又はヘッダブロック）のＩＤとは異なるＩＤが見付るまでリードする。このリードしている期間に、ヘッダブロックに対応するＣブロックをチェックし、その対応するＣブロックが見付らないヘッダブロックがあれば、それがライトバックされなかったライトコマンドである。そして、ヘッダブロックのつぎのブロックからはユーザデータが書き込まれているので、このユーザデータを読出して各ディスクドライブＤＤ０～ＤＤＮに転送してデータ修復処理を行なう。

【００４６】このようにして、小さいサイズのライトデータを各ディスクドライブＤＤ０～ＤＤＮに転送するときには、予備ディスクドライブＤＤｓをデータのバックアップ用として使用する。予備ディスクドライブＤＤｓでは、ライトデータ（ユーザデータ）の先頭にヘッダブロックが格納されて、ライトバック処理の終了後にＣブロックが格納される。ライトデータがないときでも、ＣＰＵ３はライトコマンドを予備ディスクドライブＤＤｓに出力している。この場合には、予備ＤＭＡＣ１１ｓは、ＣＰＵ３が生成したヌルブロックを繰り返し転送する。

【００４７】ホストＣＰＵ２からバックアップすべきライトデータが転送されて、キャッシュメモリ６に格納されると、ＣＰＵ３はヘッダブロックを生成して、予備ＤＭＡＣ１１ｓを介して予備ディスクドライブＤＤｓに転送する。この後に、キャッシュメモリ６に格納されたライトデータを予備ディスクドライブＤＤｓに転送し、再度ヌルブロックの転送に戻る。そして、ライトバック処理が終了すると、ＣＰＵ３はＣブロックを生成して、予備ＤＭＡＣ１１ｓを介して予備ディスクドライブＤＤｓに転送する。

【００４８】このような予備ディスクドライブＤＤｓを使用した方式であれば、予備ディスクドライブＤＤｓを使用したバックアップ処理では、シーケンシャル・ライトアクセスであるため、ドライブＤＤｓのヘッドのシーク時間およびディスクの回転待ち時間はほぼ０である。この予備ディスクドライブＤＤｓを使用したバックアップ方式とライトキャッシュ方式を併用すれば、ライトキャッシュ方式による高速アクセスとバックアップ方式によるデータの保護機能を実現することができる。したがって、結果的には二重化ライトキャッシュ付き無停電電源装置を備えたディスク・アレイ装置と同等の性能と信頼性を、二重化なしライトキャッシュ付き無停電電源装置なしの装置とほぼ同等のコストにより得ることができる。

【００４９】なお、予備ディスクドライブＤＤｓに故障が発生した場合には、他のディスクドライブＤＤ０～ＤＤＮの故障時と同様に、コントローラ１はホストＣＰＵ２またはオペレータに故障発生のアラームを発生する。このとき、ＣＰＵ３はキャッシュメモリ６の内容を全てライトバックし、予備ディスクドライブＤＤｓの修理または交換が終了するまでキャッシュメモリ６を使用したライトキャッシュ動作を停止する。

【００５０】また、ライトバック処理が終了する前に、装置の電源がオフされた場合に備えて、電源がオンされた直後にＣＰＵ３は予備ディスクドライブＤＤｓのヘッダブロックとＣブロックチェックを行なうようにしてもよい。但し、ライトバック処理が終了する前に、キャッシュメモリ６と予備ディスクドライブＤＤｓの両方が同時に故障する確率はほとんど０と考えられる。

【００５１】

【発明の効果】以上詳述したように本発明によれば、ライトキャッシュ方式を採用したディスク・アレイ装置において、予備ディスクドライブをライトデータのバックアップ用に利用することにより、キャッシュメモリやディスク・アレイ・コントローラの２重化、またはキャッシュメモリに不揮発性メモリや無停電電源装置を使用するなどの方法を採用することなく、小さいサイズのデータのライトアクセスの高速性と確実なデータ保護を実現することができる。したがって、コストの増大化やシステム構成の複雑化を招くことなく、特に小さいサイズのデータのライトアクセスの高速かつ効率化を図ると共に、データを確実に保護して装置の高信頼性を実現できることになる。

【図面の簡単な説明】

【図１】本発明の実施例に係わるディスク・アレイ装置の要部を示すブロック図。

【図２】同実施例の動作を説明するためのフローチャート。

【図３】同実施例の動作を説明するためのフローチャート。

【図４】同実施例の動作を説明するためのフローチャート。

【図５】同実施例に係わる予備ディスクドライブの記憶内容を説明するための概念図。

【図６】同実施例に係わる予備ディスクドライブの記憶内容を説明するための概念図。

13

【図7】同実施例の動作を説明するためのフローチャート。

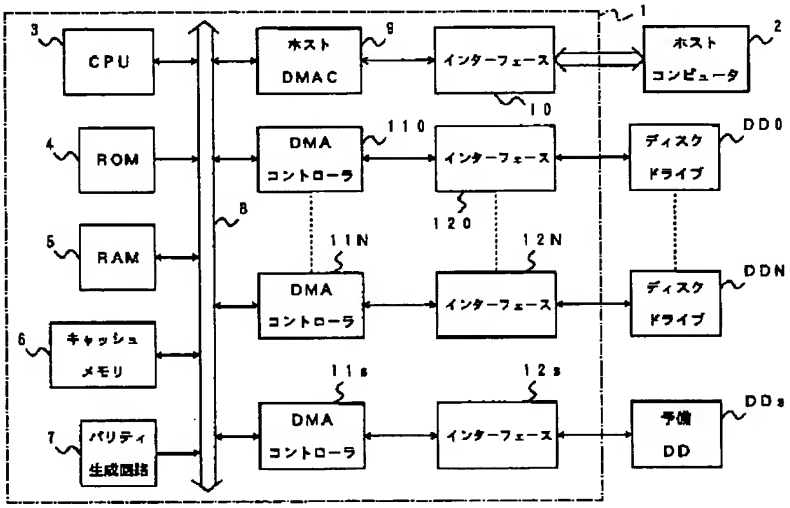
【図8】従来のレベル5のディスク・アレイ装置の構成を説明するための概念図。

【符号の説明】

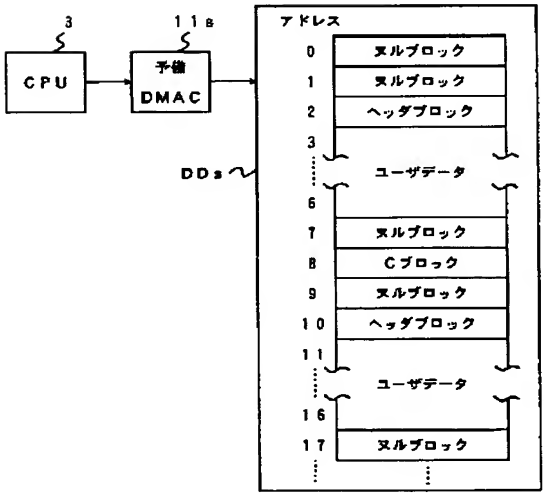
1…ディスク・アレイ・コントローラ、2…ホストコンピュータ、3…CPU、6…キャッシュメモリ、7…パリティ生成回路、9…ホストDMAコントローラ、DD0～DDN…ディスクドライブ、110～11N…ディスクDMAコントローラ、DDs…予備ディスクドライブ、11s…予備ディスクDMAコントローラ。

14

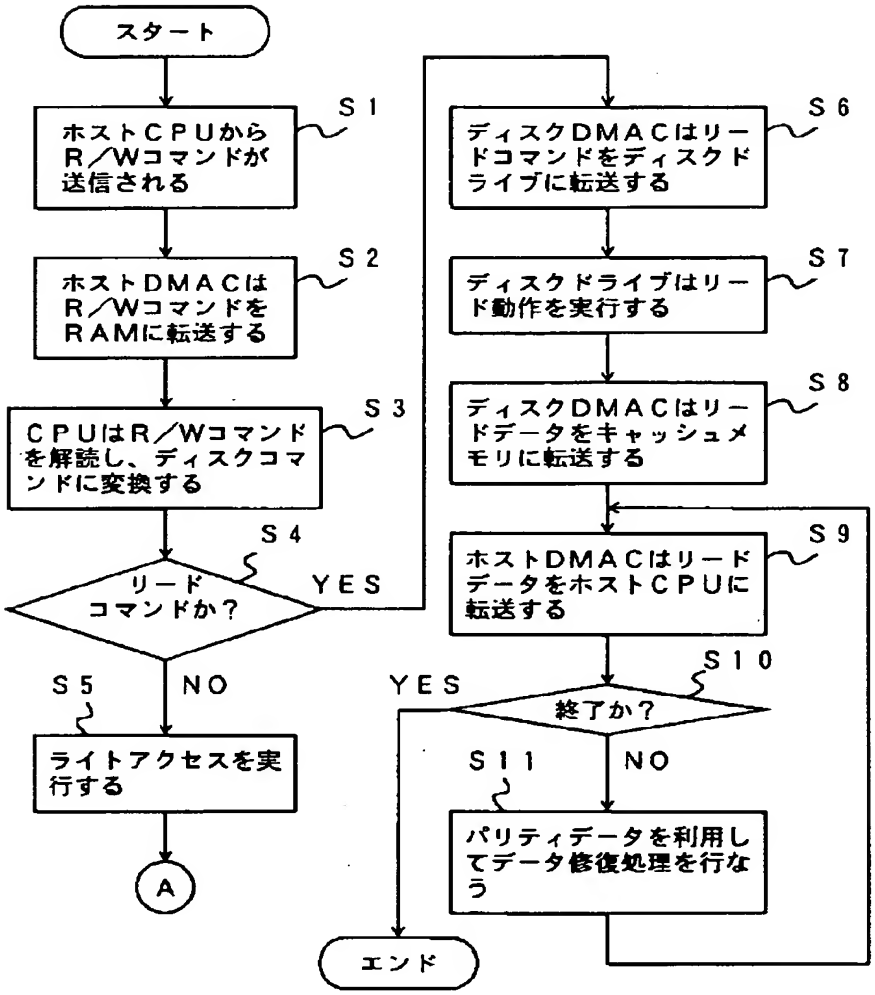
【図1】



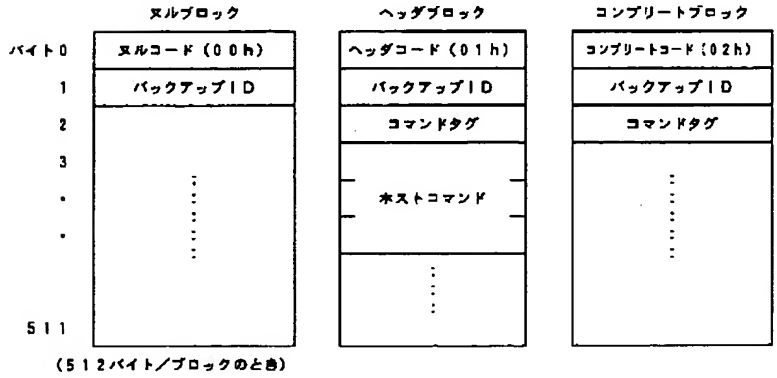
【図5】



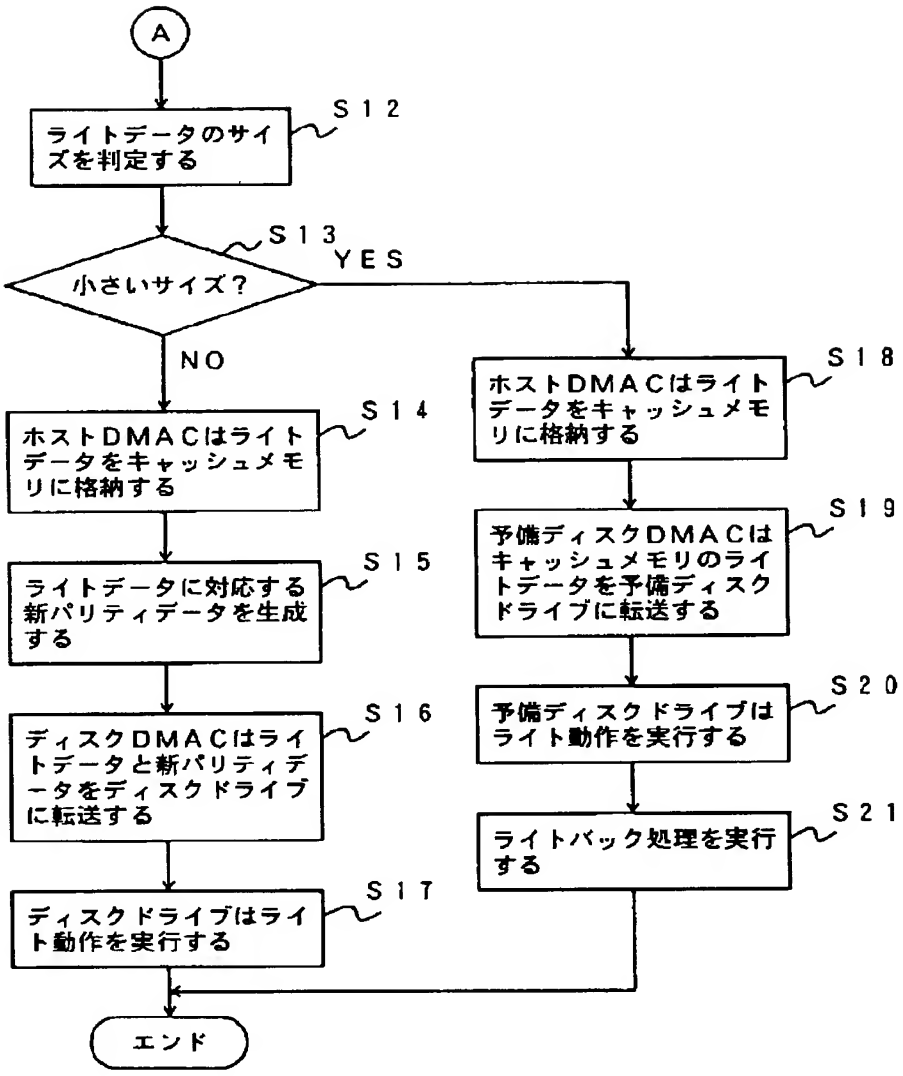
【図2】


スタート
ホストCPUからR／Wコマンドが送信される
S1
ホストDMACはR／WコマンドをRAMに転送する
S2
CPUはR／Wコマンドを解読し、ディスクコマンドに変換する
S3
S4
リードコマンドか？
YES
NO
S5
ライトアクセスを実行する
A
ディスクDMACはリードコマンドをディスクドライブに転送する
S6
ディスクドライブはリード動作を実行する
S7
ディスクDMACはリードデータをキャッシュメモリに転送する
S8
ホストDMACはリードデータをホストCPUに転送する
S9
S10
終了か？
YES
NO
S11
パリティデータを利用してデータ修復処理を行なう
エンド

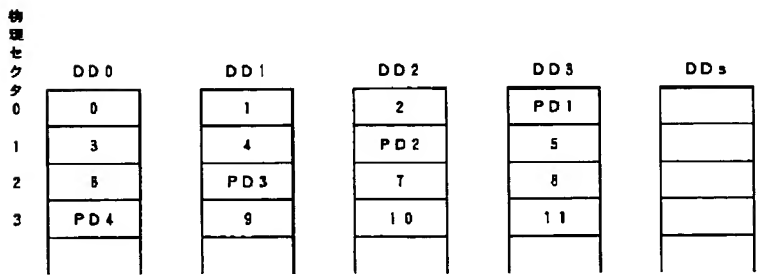
【図6】


ヌルブロック
ヘッダブロック
コンプリートブロック
バイト0
1
2
3
511
ヌルコード（00h）
バックアップID
ヘッダコード（01h）
バックアップID
コマンドタグ
ホストコマンド
コンプリートコード（02h）
バックアップID
コマンドタグ
（512バイト／ブロックのとき）

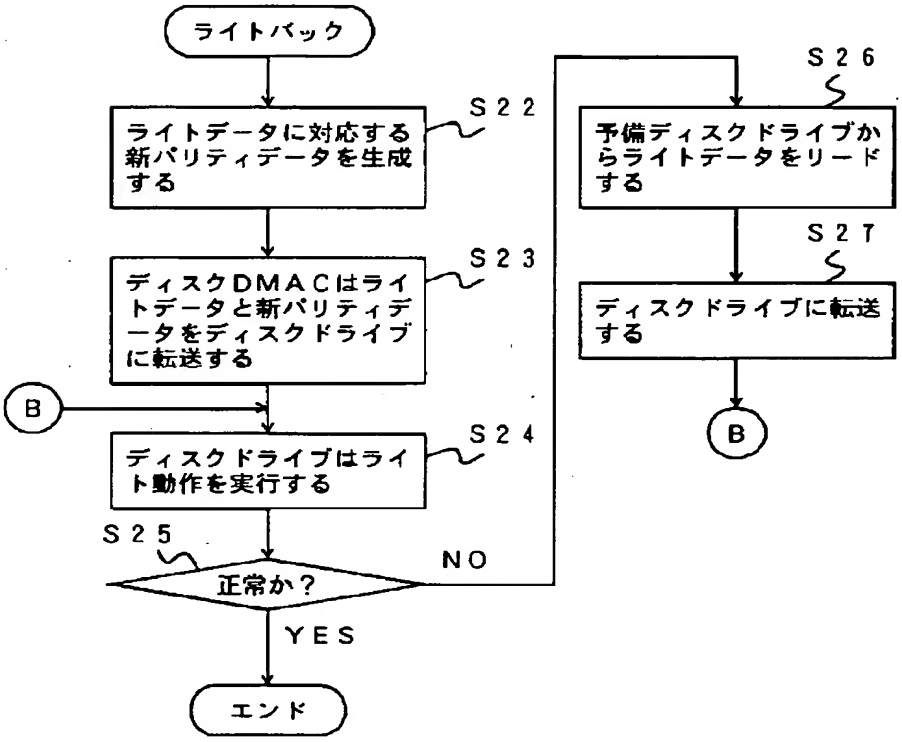
【図３】


A
ライトデータのサイズを判定する
S12
S13
小さいサイズ？
YES
NO
ホストDMACはライトデータをキャッシュメモリに格納する
S14
ライトデータに対応する新パリティデータを生成する
S15
ディスクDMACはライトデータと新パリティデータをディスクドライブに転送する
S16
ディスクドライブはライト動作を実行する
S17
ホストDMACはライトデータをキャッシュメモリに格納する
S18
予備ディスクDMACはキャッシュメモリのライトデータを予備ディスクドライブに転送する
S19
予備ディスクドライブはライト動作を実行する
S20
ライトバック処理を実行する
S21
エンド

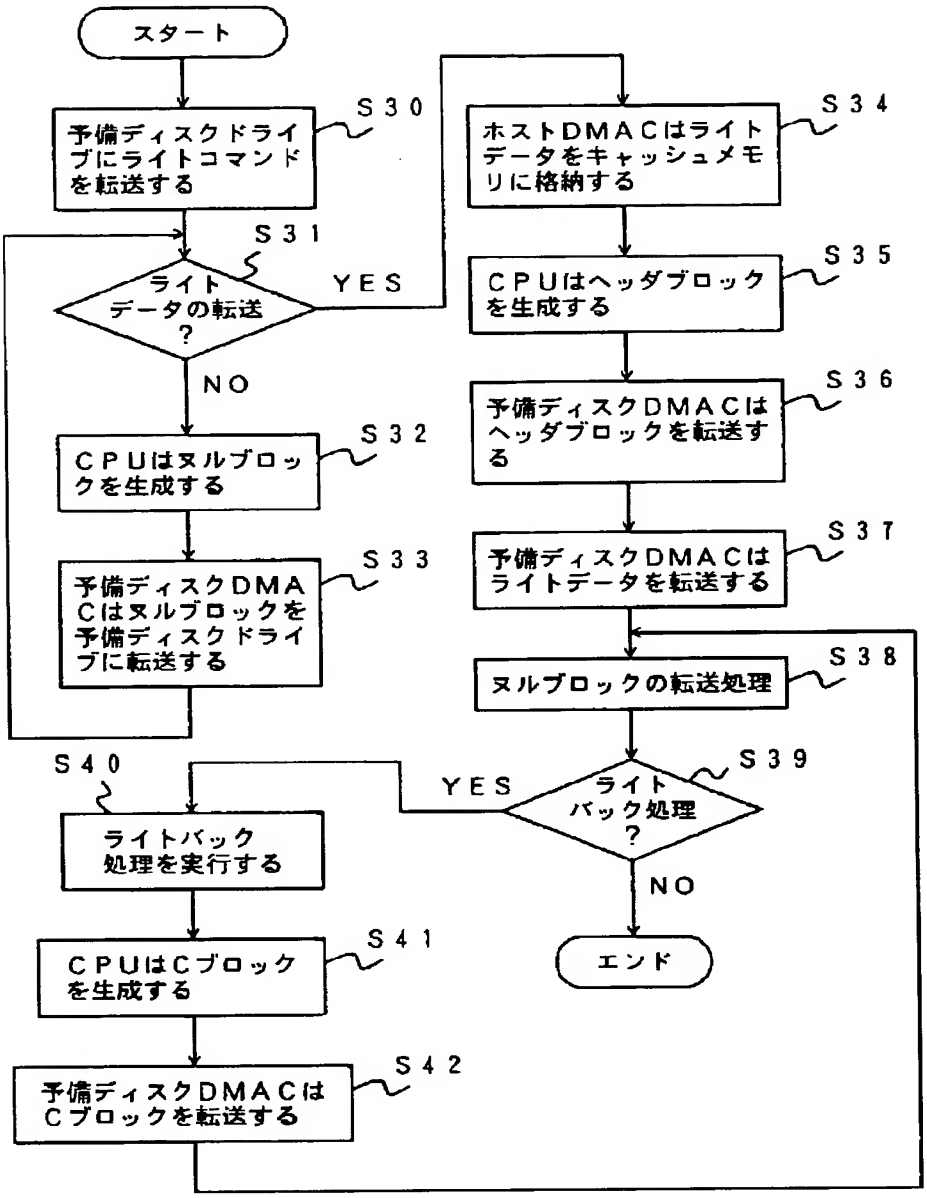
【図８】


物理セクタ
0
1
2
3
DD0
0
3
6
PD4
DD1
1
4
PD3
9
DD2
2
PD2
7
10
DD3
PD1
5
8
11
DDs

【図４】


ライトバック
ライトデータに対応する新パリティデータを生成する
S22
ディスクDMACはライトデータと新パリティデータをディスクドライブに転送する
S23
B
ディスクドライブはライト動作を実行する
S24
S25
正常か？
NO
YES
エンド
S26
予備ディスクドライブからライトデータをリードする
S27
ディスクドライブに転送する
B

【図７】


スタート
予備ディスクドライブにライトコマンドを転送する
S30
S31
ライトデータの転送？
YES
NO
CPUはヌルブロックを生成する
S32
予備ディスクDMACはヌルブロックを予備ディスクドライブに転送する
S33
ホストDMACはライトデータをキャッシュメモリに格納する
S34
CPUはヘッダブロックを生成する
S35
予備ディスクDMACはヘッダブロックを転送する
S36
予備ディスクDMACはライトデータを転送する
S37
ヌルブロックの転送処理
S38
ライトバック処理？
S39
YES
NO
S40
ライトバック処理を実行する
CPUはCブロックを生成する
S41
予備ディスクDMACはCブロックを転送する
S42
エンド